Estimating Energy Intake of Urban Women in Colombia: Comparison of Diet Records and Recalls

DARNA L. DUFOUR,[1,5]* LISA K. STATEN,[2] CAROL I. WASLIEN,[3] JULIO C. REINA,[4] AND G.B. SPURR[5,6,7]
[1]*Department of Anthropology, University of Colorado, Boulder, Colorado 80309–0233*
[2]*Department of Family and Community Medicine, University of Arizona, Tucson, Arizona 85712*
[3]*University of Hawaii School of Public Health, Honolulu, Hawaii 96822*
[4]*Department of Pediatrics, Universidad del Valle, Cali, Colombia*
[5]*Department of Physiological Sciences, Universidad del Valle, Cali, Colombia*
[6]*Department of Physiology, Medical College of Wisconsin, Milwaukee, Wisconsin 53226*
[7]*Research Service, Zablocki Veterans' Administration Center, Milwaukee, Wisconsin 53226*

KEY WORDS nutrition; methods; 24 h recalls; estimated records; food groups; developing country

ABSTRACT As part of a larger study of energy-nutrition, we compared the performance of 24 h diet recalls with estimated diet records kept by trained observers. The subjects were economically disadvantaged women (n = 85) in the city of Cali, Colombia. A 24 h recall and an estimated diet record were collected for each woman at 0 and approximately 3 and 6 months. Energy intake obtained from the estimated dietary records was validated against energy expenditure and used as the reference method. Energy and macronutrient intake were calculated from published food composition tables and proximate analyses of common foods. The number of food items consumed per woman per day, total and in each of 16 food groups, was tabulated. Energy and macronutrient intakes were 11–13% lower in the 24 h recalls. The discrepancy energy could be largely accounted for by the lower number of food items in the recalls. The number of food items in eight of 16 food groups was significantly lower in the recalls compared to the records. Underreporting on the recalls was a general tendency in these subjects and not clearly related to average energy intake. We conclude that 24 h diet recalls underestimate energy and nutrient intake in this population and are not suitable for studies of human energetics. Am J Phys Anthropol 108:53–63, 1999. © 1999 Wiley-Liss, Inc.

Estimating the diet intake of a free-living population is important in addressing many anthropological questions of nutritional adaptation, energetics, and epidemiology (Huss-Ashmore, 1996), but it is a difficult task. There are a number of methods available, each of which has advantages and disadvantages (Bingham and Nelson, 1991). The 24 h recall method, in which the subject is asked to recall all food eaten in the past 24 h, has probably been the most widely used in field situations. It is a method that does not depend on subject literacy and has a number of advantages in terms of ease of administration and low cost. A major disadvantage is its dependence on the recall ability of subjects.

Grant sponsor: NIH; Grant number: 5-R22-DK39734.
*Correspondence to: Darna L. Dufour, Anthropology Department, University of Colorado, Boulder, CO 80309–0233. E-mail: dufourd@spot.Colorado.edu
Received 26 February 1998; accepted 4 October 1998.

© 1999 WILEY-LISS, INC.

A number of studies have shown that 24 h recall data tends to underestimate nutrient intake obtained by weighed records, which are generally considered the gold standard (Bingham, 1987). Other studies, however, have shown that 24 h recalls can yield similar and even higher nutrient intakes (Blake and Durnin, 1963; Greger and Etnyer, 1978; Ohlson et al., 1950). In trying to understand these differences, most researchers have focused on problems associated with the estimation of food portion size. However, the first thing a subject has to do in completing a 24 h recall is remember that a particular food item was consumed. Conceptualizing the size of the food item and effectively communicating that information to the interviewer are the second and third steps, respectively. Although errors are possible in each or all three of these steps, surprisingly few studies (Acheson et al., 1980; Linusson et al., 1974; Thomson 1958) have considered the accuracy of the first step, the recall of food items, upon which everything else depends.

In addition to problems of memory, reports of food intake obtained using the 24 h recall protocol may be biased in several other ways. For example, subjects may not accurately report intake for reasons related to the social desirability of consuming specific foods (Thompson and Byers, 1994) or population specific conceptions of how much one should eat. With regard to the latter, two studies in the USA have reported a tendency for individuals with relatively high intakes on diet records to underreport their intake on 24 h recalls and visa versa (Linusson et al., 1974; Maddan et al., 1976; Gersovitz et al., 1978). This phenomenon is known as the flattened slope syndrome since regressions of intake obtained by recall on intake obtained by records typically exhibit slopes of less than one. Whether this phenomenon occurs in other populations is not known.

The objective of this paper is to evaluate the information on diet intake obtained using 24 h recalls in a population of low-income urban women living under conditions of economic disadvantage in Cali, Colombia. In evaluating the recall data, we used the estimated diet records as a reference since diet records are assumed to involve fewer sources of error than recalls (Bingham and Nelson, 1991), and energy intake obtained from estimated records was validated against energy expenditure (Spurr et al., 1996). In this paper we test three hypotheses: 1) that energy and macronutrient intake obtained by recall is lower than that obtained from diet records, as is the number of food items; 2) that all foods are recalled with the same degree of accuracy; and 3) that there is no difference in degree of underreporting on recalls in women with low-average vs. high-average intakes.

The research reported here was done as part of a larger study of energy nutrition of urban women living in poor neighborhoods of Cali, Colombia. One objective of the larger study was to use energy intake as an indicator of undernutrition. Because of the low level of literacy of some of the subjects, we decided to rely on food records estimated by trained observers. A 24 h recall protocol was included because we were unable to obtain diet records covering weekend days and were interested in knowing if there were weekday-weekend differences in energy and macronutrient intake.

SUBJECTS AND METHODS

Subjects and setting

The subjects were 85 nonpregnant, nonlactating women (aged 19–43 years), all of the women were volunteers and had been recruited through word of mouth. All subjects received a physical exam by a physician before being admitted into the study. Informed consent was secured from all subjects. The protocols were approved by the Human Research Review Committee of the Medical College of Wisconsin and the Research Committee of the Universidad del Valle.

The women lived in poor neighborhoods on the periphery of Cali, a city of ~2.2 million people. Living conditions are described in Dufour et al. (1997). The women purchased most food at the small stores in their neighborhood. Some women purchased dry goods (grains, legumes, sugar, coffee, etc.) weekly and fresh products (bread, meat, fruits, vegetables) daily, and others purchased all food daily or by the meal. The subjects themselves did most of the food

preparation in their own households and ate most meals at home.

Study design

Diet intake was assessed in three measurement rounds at 0 and approximately 3 and 6 months. Each measurement round lasted 1 week, during which diet intake was assessed for 4 different days. Estimated diet records were used on 2 consecutive weekdays, and 24 h recalls were used on 2 other days, the second of which was always a weekday. Other than these constraints, the specific days of dietary assessment were assigned randomly to individual subjects. In this paper, we compare mean values obtained from 1 day of record (day 2) and 1 day of recall (day 2) from each measure round. The second day of records and recalls in each round was used to eliminate potential weekend-weekday differences and to minimize any observer effect. Data were collected between August 1990 and May 1995. Days of food intake were representative of normal intake as far as we were able to determine.

Anthropometry and sociodemographic characteristics

Measurements of body weight and height were taken by a trained technician following standard techniques described by Lohman et al. (1988). Body weight was measured in kilograms (Homs Beam Balance, Douglas Homs Corp., Belmont, CA) (±25 g) with the subjects lightly clothed and without shoes. Height was measured in meters using a wall stadiometer. Body mass index (BMI) was calculated as weight/height^2. Sociodemographic information was obtained by structured interview and observation.

Observer training and accuracy in estimating portion size

All dietary data were collected by trained observers. The observers were young women (aged 18–25 years) who lived in the same neighborhoods as the subjects and were trained to unobtrusively record both physical activity and diet intake. They worked in pairs and alternated 4 h shifts scheduled to cover the awake portion of subjects' days (about 16 h).

Training occurred over a 5 week period and involved 1) the development of consensus on the names of foods and food preparations, names of common serving utensils, sizes of food items (small, medium, large, etc.), and descriptions of the fullness of serving utensils (half-full, full, heaping, etc.); 2) numerous practice sessions in sizing food items and estimating fullness of serving utensils; and 3) practice in the collection of weighed recipes from local women. As observer competence advanced, training incorporated the development of a data table containing average weights and edible portions of common food items. The table was developed by purchasing representative samples of foods in local shops and weighing them in the laboratory, and an average weight was established for each type of food and size category. Breads were sized by monetary units (i.e., a 50 peso bread), as that is the way they are named. Since bread size tended to change over time, a representative sample of breads was obtained from vendors every 6–12 months and weighed in the laboratory. For foods served using spoons or cups, the conversion of serving sizes to gram weights was based on an extensive series of repeat measurements of common foods done in the laboratory using a variety of common household utensils. All foods were weighed on a Sunbeam electronic balance (Oster, Mcminnvile, TN) (±3 g) calibrated periodically with standard weights.

The accuracy of the observers in estimating serving sizes was assessed periodically by the following procedure. Local women who had participated in the project were visited at meal time and asked to serve plates of food as they normally did. One of us (D.L.D.) weighed the plates on a dietary balance (Sunbeam electronic, ±3 g) as each food was added. At the same time, the observers recorded the serving size of each food item in household measures. The scale weights were not visible to the observers at any time. The estimated weights were later converted to metric weights as described below.

Estimated records

In recording food intake, observers estimated the serving size of food items as they were served to or selected by the subject. For foods served by the piece, such as pieces of

meat or fruit, the observers recorded a description of the food and its relative size (small, medium, large). For foods served using spoons or cups, observers recorded the type of spoon or cup used and the degree of fullness. Descriptions of food item sizes were later converted to gram weights using the food portion size data table described above. In addition, volumetric measurements of serving utensils and drinking vessels were completed in each subject's home before observations of the food intake for that subject began, and these measurements were used as appropriate. Observers were also trained to obtain serving size weights of all new and unusual foods and common foods served with unusual utensils and to add this data to the portion size table as the study progressed.

Twenty-four hour recalls

The same trained observers responsible for the diet records also administered the 24 h recalls. The diet records and 24 h recalls were done during the same week in each measurement round but covered different days. Estimated food quantities were converted to gram weights as described above.

Calculation of nutrient intake

Energy and nutrient composition of foods were obtained from published food composition tables (Instituto Nacional de Nutrición, 1988; Pennington and Church, 1985; USDA, 1994) and proximate analyses of samples of the most commonly consumed foods. The latter were completed by the Instituto Nacional de Nutrición in Bogotá. Recipes (in household units) were collected for composite dishes and beverages. For the more common dishes and beverages, weighed recipes were collected, the food was prepared in the laboratory to determine the final cooked weight of the recipe, and nutrient composition was calculated from published tables or proximate analyses. For packaged snack foods, nutrient composition was obtained from information on the package. Energy and nutrient intakes were calculated for all intake data using custom software programs written by G.B.S.

Types of foods consumed

The food items listed in both the records and recalls were aggregated into 16 food groups on the basis of shared characteristics. The food groups are as follows:

1. Alcohol
2. Bread, pastry, cookies
3. Candy and other sweets
4. Coffee, hot chocolate, aguapanela (a sugar-water drink)
5. Fruit juices (made with water or milk, and added sugar)
6. Fruit
7. Legumes
8. Meat, fish, offal, canned meats and fish, mixed dishes with meat or fish
9. Milk and milk products
10. Mixed vegetable dishes
11. Other (condiments, added fats, uncommon foods)
12. Rice, rice-based dishes, pasta, pasta-based dishes
13. Roots, tubers, plantains
14. Salads, cooked vegetables
15. Soft drinks, other drinks (powdered drink mixes, starch thickened drinks, corn-based drinks)
16. Soup, stew

In the following text, these groups are referred to by the first item, which was the most frequently consumed item in the group.

Number of food items

The number of food items in the records and recalls, total and subtotal for each food group, was tabulated and the average number of items per woman per day calculated. A food item was defined as a food recorded or recalled as being consumed as a single unit or serving at a given time. For example, a serving of rice (made up of one or more spoonfuls) at a given meal would be recorded as a single food item, a glass of juice as a second food item. Another serving of rice eaten at the same meal, or later the same day, would be recorded as a third food item.

Data analysis

Random samples of data were checked periodically for coding errors. For all subjects with unusually high or low energy

intakes, the original data sheets were checked for possible coding errors and handwritten comments on the subject that might clarify unusual values.

To evaluate the accuracy of energy values calculated from observer estimated weights, we computed Pearson correlation coefficients between each observer and the scale weight and among the three observers. To control for Type I error across six correlations, we used the Bonferroni procedure, and a P value of <0.008 (equal to 0.05 divided by the number of correlations [n = 6]) was considered significant. Agreement between energy values calculated from estimated and actual food weights was also evaluated graphically using the method of Bland and Altman (1986), in which the differences between the values obtained with different methods are plotted against their mean. A similar analysis was done for the actual and estimated food weights themselves.

Two-way within-subject analysis of variance (ANOVA) with repeat measures was used to evaluate the effect of method and time on macronutrient intake. The dependent variable was intake in kilojoules or grams; the within subjects factors were methods (record vs. recall) and time (three levels [i.e., rounds 1, 2, 3]). Wilks's lambda (Λ) was used as the criterion of significance. Paired *t*-tests were used to compare the mean number of food items/woman/day in the diet records and recall and the mean number of food items in each of the 16 food groups in the records and recalls. The Bonferroni procedure was used to control for Type I error across multiple tests, and a P value of .0045 (.05/11) was considered significant. Linear regression analysis was used to evaluate the prediction of total energy intake from the number of food items/woman/day. Pearson correlation coefficients were calculated to evaluate the linear relationship between sociodemographic characteristics (age and education) and the magnitude of the discrepancy between the methods.

To test the hypothesis that subjects with low-average energy intakes will tend to overreport on the recall and that subjects with high-average intakes will tend to underreport intake on the recalls, we divided the subjects into tertiles based on mean energy intake in the records. The difference between energy intake in the records (mean of 3 days) and recalls (mean of 3 days) was calculated for each subject. Mean values for 3 days were used to reduce some of the day-to-day variability in intake. The pooled mean energy intakes of the three tertiles were compared using one-way ANOVA.

All statistical analyses were completed using SPSS 7.5 (SPSS Inc., Chicago IL), and significance was set at $P = 0.05$, except where noted.

TABLE 1. Correlation among energy values calculated from scale weights and weights estimated by three observers for n = 38 food items

	Scale	Observer 1	Observer 2	Observer 3
Scale	1.00			
Observer 1	.856*	1.00		
Observer 2	.854*	.934*	1.00	
Observer 3	.894*	.956*	.897	1.00

* $P < .008$.

RESULTS

Observer accuracy in estimating food weights

The energy values calculated from actual weights of foods were compared with those calculated from estimated weights for 38 food items served at midday meals in different households. The food items were rice (n = 12), pasta (n = 5), beans (n = 3), soup and stew (n = 7), meat (n = 4), fried plantain slices (n = 4), and salad (n = 4). The correlation between the energy values calculated from scale weights and estimated weights was high (r = 0.82–0.88, $P < 0.008$), as were correlation coefficients between observers (Table 1). The total energy value of all 38 food items calculated from scale weights was not significantly different from that calculated from the estimated weights (paired *t*-test, $t = -0.21$, $P = 0.837$). Coefficients of variation (CV) were also calculated for the scale and estimated weights (mean of three observers) for each of the foods. These values ranged from 0.4–42%, with a mean of 14 ± 11.4%.

Results of the Bland-Altman analysis are shown as two scatterplots, one of the weights of food items (Fig. 1A) and the other of the

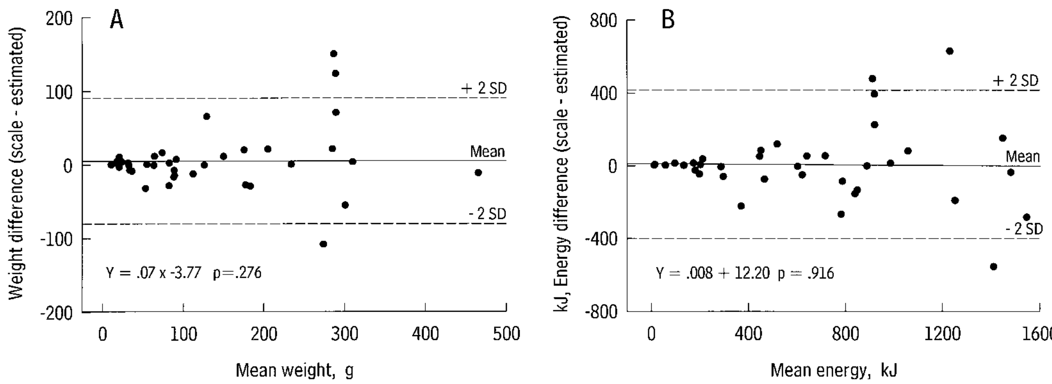

Fig. 1. **A:** Scatterplot of differences between scale weight and estimated weight and mean of two measurements following method of Bland and Altman (1986). **B:** Scatterplot of differences in energy values calculated from scale weight and estimated weight and energy values calculated from the mean of the scale and estimated weights following method of Bland and Altman (1986).

energy values of the same food items (Fig. 1B). For the weights of food items, overall agreement between methods is good. For the food items (n = 8) with serving sizes greater than 250 g, there is more scatter, suggesting lower observer accuracy (Fig. 1A). The three food items with the lowest accuracy were soup (n = 1) and stew (n = 2). Energy values calculated from actual and estimated food weights show a similar pattern of distribution (Fig. 1B). The regression of weight differences (scale weight minus estimated weight) on mean weight (Fig. 1A) has a slope which is not significantly different from zero, indicating that the size of the serving has no effect on the difference between the two methods. Similarly, when the values are expressed in terms of energy, the slope of the regression is not significant. The mean difference in energy between the two methods (scale weight minus estimated weight) was 6.94 ± 204.33 kJ (1.66 ± 48.84 kcal), which is equal to .03% of the total energy value of the weighed foods.

Anthropometric and sociodemographic characteristics of subjects

Mean stature and body weight of the subjects were 155.6 ± 6.1 cm and 53.9 ± 7.3 kg, respectively. Mean BMI was within the normal range, 22.2 ± 2.6 kg/m^2. The subjects ranged in age from 19–43 years, with a mean of 29.0 ± 6.2 years. Formal education averaged 6.3 ± 2.7 years (Table 2).

TABLE 2. Sociodemographic characteristics of Cali women

Group age (years)	n	Education completed (years) Mean ± SD	Range
19	3	7.3 ± 0.8	7–8
20–24	23	7.3 ± 2.5	3–13
25–29	17	6.5 ± 2.3	2–10
30–34	21	6.6 ± 3.4	0–14
35–39	19	4.8 ± 1.9	2–9
40–43	2	4.5 ± 2.1	3–6
Total	85	6.3 ± 2.7	0–14

Comparison of records and recalls

Energy and macronutrient intakes obtained using records and recalls on weekdays are shown in Table 3. In the records, mean intakes of energy, carbohydrate, protein, and fat were 8,814 ± 1,742 kJ (2,106.6 ± 316.4 kcal), 372.2 ± 74.32 g, 61.0 ± 14.01 g, and 45.4 ± 13.72g per day, respectively. These values were significantly greater than those obtained using 24 h recalls (two-way ANOVA with repeat measures, (= 0.80, F(1.84) = 20.49, $P < 0.001$). Absolute values obtained by records were higher by 11–13% than those obtained by recall. The time effect in the ANOVA model was not significant.

The average number of food items/women/day was significantly higher in the records than in the recalls, 15.4 ± 4.1 as opposed to 13.1 (± 4.1 food items (t = -8.142, $P < 0.001$).

TABLE 3. Energy and macronutrient intake of Cali women (n = 85) from diet records (3 nonconsecutive days) and 24 h recalls (3 nonconsecutive days)

Nutrient	Records (mean ± SD)	24 h recalls (mean ± SD)	Significance
Energy (kJ)	8,814 ± 1,742	7,923 ± 2,024	<.001
Energy (kcal)	2,107 ± 316.4	1,894 ± 483.9	
Carbohydrate (g)	372.2 ± 74.32	334.2 ± 82.00	<.001
Protein (g)	61.0 ± 14.01	54.0 ± 15.79	<.001
Fat (g)	45.4 ± 13.72	40.6 ± 15.82	.006

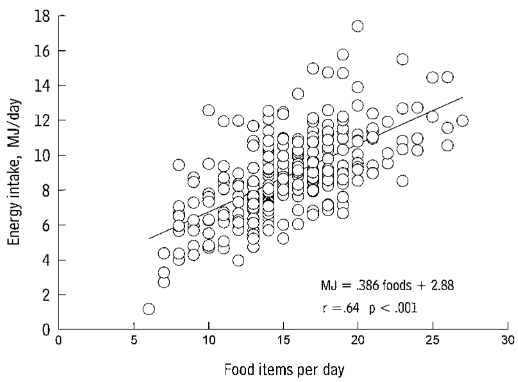

Fig. 2. Scatterplot of energy intake (MJ/day) against number of food items/woman/day in diet records for 85 women for 3 days each.

The discrepancy between records and recalls was not correlated with either age (r = -.008, $P = 0.467$) or years of education (r = -0.19, $P = 0.08$). The linear relationship between energy intake (records) and number of food items food items/woman/day is shown in the scatterplot in Figure 2. The regression equation for predicting energy intake from food items/woman/day is energy intake, MJ = 0.386 · food items + 2.88. The number of food items/woman/day accounts for approximately 41% of the variance in energy intake. The correlation between energy intake and items/woman/day is 0.64 ($P < 0.001$).

Types of foods in records and recalls

The average number of food items/woman/3 intake days in each of 16 food groups is compared for records and recalls in Table 4. The four most frequently consumed foods in both the records and recalls were coffee, rice, bread, and meat. These four food groups accounted for over 50% of the total dietary energy intake in the records. The average number of food items in each food group tended to be higher in the records and was significantly higher in seven of 16 food groups. In only one food group, soft drinks, were the recall values higher than the records. The food groups showing the greatest discrepancies between the two methods were coffee, bread, meat, fruit, and salad. Food groups showing relatively good agreement between the records and recalls were rice, roots, fruit juices, legumes, candy, mixed vegetable dishes, milk, and other.

Is underreporting on recalls related to average energy intake?

The difference between mean energy intake in records and recalls for each subject is plotted by tertile of intake in Figure 3. For subjects in the middle tertile, the average intake group, most of the differences between methods (records minus recalls) were positive, indicating a tendency to underreporting. Subjects in the high intake tertile show a similar pattern. The mean differences between methods are positive, 1.1 ± 1.37 and 1.4 ± 1.89 in the middle and high energy intake tertiles, respectively. For subjects in the low-intake tertile, the mean difference is also positive but absolutely smaller, 0.12 ± 1.92 MJ, indicating that on average there is less discrepancy between recalls and records. There is, however, more variability in the low-intake tertile. The means of the low-intake and high-intake tertiles are not significantly different from that of the middle tertile, but they are significantly different from each other (one-way ANOVA, Bonferroni post hoc test, $P = 0.017$). A similar analysis of the number of food items in the records and recalls for each tertile gave similar results.

TABLE 4. Comparison of mean number food items per woman (n = 85) for 3 days of intake by food group in diet records and 24 h recalls[1]

Food group	Items/woman/3 intake days		
	Record	Recall	Significance[2]
1. Coffee, chocolate, aguapanela	9.0 ± 5.12	7.1 ± 3.86	<.001
2. Rice, rice dishes, pasta, pasta dishes	6.0 ± 1.87	5.5 ± 1.64	.014
3. Bread, pastry, cookies	5.7 ± 2.24	4.7 ± 2.47	<.001
4. Meat, fish and mixed dishes	5.7 ± 2.1	4.9 ± 2.00	<.001
5. Roots, tubers, plantains	3.0 ± 2.14	2.4 ± 1.61	.028
6. Soft drinks, other drinks	2.9 ± 2.00	4.9 ± 1.93	<.001
7. Fruit juices	2.4 ± 1.93	2.1 ± 2.00	.204
8. Fruits	2.0 ± 2.44	0.9 ± 1.29	<.001
9. Soup, stew	1.9 ± 1.44	1.4 ± 1.43	.003
10. Legumes	1.8 ± 1.66	1.5 ± 1.44	.166
11. Salads, vegetables	1.2 ± 1.34	0.6 ± 0.98	.001
12. Other	1.2 ± 1.48	0.7 ± 0.89	.004
13. Candy	1.1 ± 1.40	0.9 ± 1.34	.320
14. Mixed vegetable dishes	0.7 ± 1.00	0.4 ± 0.75	.068
15. Milk and milk products	0.7 ± 1.07	0.5 ± 1.03	.249
16. Alcohol	0.1 ± 0.40	0.1 ± 0.38	.436

[1] Food groups listed by frequency of consumption in the diet records.
[2] $P < .0045$ considered significant.

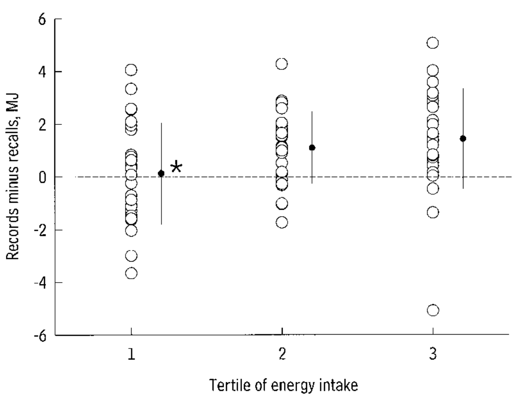

Fig. 3. Scatterplot of differences between records and recalls plotted by tertile of mean energy intake in records (1 = low-intake, 3 = high-intake). Values are means of records (n = 85) and means of recalls (n = 85). Mean ± SD are shown for each tertile. *Significantly different than tertile 3 (one-way ANOVA, Bonferroni post hoc comparisons, $P = 0.017$)

DISCUSSION

Observer accuracy

The accuracy of observers in estimating weights of common food items was good, and energy values calculated with estimated weights were not significantly different than those calculated from scale weights. Both over- and underestimation occurred but tended to cancel one another out such that the mean difference between actual and estimated weights was low. The greatest discrepancies between actual and estimated weights were for food items (soups and stews) with total portion sizes greater than 250 g. The lower accuracy of observers in estimating the weight of food served in multiple spoonfuls results probably because each spoonful may have a different level of fullness, and the repeat spoonfuls typically occur in rapid succession. The greater variability in the estimated weights of these types of foods probably had a minimal impact on the calculation of daily energy intake because food items with weights of 250 g or greater occurred only in 20% of all food items in the diet records.

A recent study by Romieu et al. (1997) also reported good agreement between actual food weights and weights estimated by trained observers. We concur with Romieu and colleagues that adequate training requires a substantial time investment. Other studies using trained observers have reported lower levels of agreement between actual and estimated weights, with CVs of 16–53% (Rutishauser, 1982; cited in Bingham and Nelson, 1991) and 21–70% (Bollard et al., 1988). Studies using untrained subjects have reported higher CVs (between 1 and 96%) (Young et al., 1952), as would be expected.

Records vs. recalls

For energy and macronutrient intake, the values obtained using 24 h recalls were

11–13% lower than the values obtained using records. Although an error of that magnitude might be acceptable in some studies, it was not acceptable in this one because one of our goals was to measure the components of energy balance at the group level.

The lower mean energy intake calculated from the recalls can be almost entirely explained by the lower number of food items in the recalls. The differences between records and recalls were 891 kJ for energy intake and 2.3 food items. Based on the regression equation in Figure 2, a difference of 2.3 food items is equal to 888 kJ. Only a few other studies have reported data on the extent to which the number of food items consumed is in error (Bingham, 1987). Acheson et al. (1980) found that men who recorded their own food intake and were then asked to fill out a questionnaire reporting what they had eaten in the past 24 h usually left out at least one food item. The incorrect number of food items, together with a tendency to underestimate portion size, resulted in an underestimation of energy intake by 21%. Thomson (1958) found that pregnant women in Scotland who kept weighed food records and were then asked to recall food intake for a 24 h period covered by the records also omitted food items. The 24 h recalls underestimated energy intake by 17%, approximately 5% of which could be accounted for by missing food items. Linusson et al. (1974) kept weighed diet records for lactating women (n = 88) in a hospital setting and found that women omitted food items in 24 h recalls, although which or how many is not clear. The women also underestimated food quantities in eight of 14 food groups.

In the present study, virtually all of the difference between energy intake in the records and recalls can be accounted for by a lower number of food items in the recalls, whereas in the other studies portion size estimation appeared to be more of a problem. Important differences between our study and the others might help explain these discrepancies. First, in this study trained observers did both the records and recalls and had completed a calibration of household dishes and serving utensils before administering the recall. This should have improved the accuracy of serving size estimation. Second, subjects in our study were women living in poverty for whom the acquisition of food was constrained by limited budgets, and high food intake had positive social value. Under such circumstances, it is likely that subjects are more focused on food quantities than are nondieting subjects living in food-rich environments. The women in our study impressed us with their accuracy in describing food quantities and their ability to divide large amounts of food into multiple, equal-sized servings. Further, in a cultural context in which being able to consume large amounts of food is valued, a bias toward underestimating food eaten is unlikely. Third, the observers in our study were well trained and shared the language related to food and food portions with the subjects. This should have facilitated communication of portion size from the subject and its correct coding by the observer.

It is clear that the subjects in the present study tended to recall some foods better than others (Table 3). One might assume that the food items most frequently consumed would be recalled most accurately, but this was not the case. Three of the four most frequently consumed items (coffee, bread, meat) appeared less frequently in the recalls than in the records. On the other hand, values for the four least frequently consumed foods (candy, mixed vegetable dishes, milk and alcohol) were similar in the recall and records. Other foods that appeared with comparable frequencies in the recalls and records were rice, roots, fruit juices, and legumes. These foods were typical components of the midday meal, the main meal of the day and the culturally most significant meal (Dufour et al., 1997). Meat and salad are also normally part of the midday meal, but they were relatively underreported in the recalls. Fruit, a food eaten between meals, also showed significant underreporting in the recalls relative to the records.

Other studies looking at the agreement between actual and recalled diet intake in terms of foods have also found larger discrepancies for some foods than others (Thomas et al., 1954; Adelson, 1960; Bransby et al., 1948; Linusson et al., 1974; Hankin et al., 1975). The foods showing the largest discrep-

ancies vary from study to study, suggesting population-specific factors are associated with the recall of specific foods. One population-specific factor is obviously the choice of core foods. For example, a study of adult males in the USA by Hankin et al. (1975) found that items eaten regularly (i.e., the core foods) tended to be recalled more accurately, but that generalization did not apply to all food items. It is clear, then, that knowledge of the type of food items most likely to be omitted in the recalls can be used to improve interviewer's ability to probe in the interview. In addition, knowledge of the types of food items most likely to be missed using recall techniques in a given population would also be important in epidemiological studies examining the link between disease prevalence and intake of specific foods. For example, a study of the relationship between cancer and fruit and vegetable intake in Cali women would be compromised by the significant underreporting of those foods using 24 h recall.

When the Cali subjects were grouped according to tertile of energy intake, we found a tendency to underreport on the recalls relative to the records in the middle and high energy intake tertiles. Subjects in the low energy intake tertile showed the best average agreement between records and recalls but the greatest variability. In this tertile, the tendency of some women to underreport on the recalls was balanced by overreporting by an approximately equal number of women. Hence, in this group of subjects we find no clear support for the idea that under- and overreporting on recalls is related to average energy intake measured using food records.

In conclusion, we found that 24 h recalls underestimated energy intake by 11% in comparison to estimated records kept by trained observers. The underestimation can be almost entirely explained by the fewer number of food items in the recalls. Some food items tended to be more accurately recalled than other foods. Underreporting on recalls was a general tendency in this group of subjects, and under- and overreporting on recalls was not clearly related to average energy intake.

ACKNOWLEDGMENTS

The authors express their gratitude to Zoila de Maldonado, Betty de Orozco, and Jairo Ardila for their excellent technical support and also to Blanca Diaz, Clara Ines Gil Rivas, Mayerling Valencia Castro, Rudt Garcia Ramirez, and Yasmin Villota for their invaluable assistance in data collection. We are also grateful to Coldeportes del Valle for logistical support, to the Instituto Nacional de Nutrición for the analysis of food samples, and to Fundación para la Educación Superior for their continued support. We thank the women of Agua Blanca for their patience and cooperation. Support for data analysis and manuscript preparation was provided by a CRCW faculty fellowship to D.L.D. from the University of Colorado.

LITERATURE CITED

Acheson KJ, Campbell IT, Edholm OG, Miller DS, Stock MJ. 1980. The measurement of food and energy intake in man—an evaluation of some techniques. Am J Clin Nutr 33:1147–1154.

Adelson SF. 1960. Some problems in collecting dietary data from individuals. J Am Diet Assoc 36:453–461.

Bingham SA. 1987. The dietary assessment of individuals: methods, accuracy, new techniques and recommendations. Nutrition Abstracts and Reviews (Series A) 57:705–742.

Bingham SA, Nelson M. 1991. Assessment of food consumption and nutrient intake. In: Margette BM, Nelson M, editors. Design concepts in nutritional epidemiology. New York: Oxford. p 153–191.

Blake EC, Durnin JVGA. 1963. Dietary values for a 24 hour recall compared to a 7-day survey on elderly people. Proc Nutr Soc 22:i.

Bland JM, Altman DG. 1986. Statistical methods for assessing agreement between two methods of clinical measurement, Lancet 1:307–310.

Bollard JE, Yuhas JA, Bollard TW. 1988. Estimation of food portion sizes: effectiveness of training. J Am Diet Assoc 88:817–820.

Bransby ER, Daubney CG, King J. 1948. Comparison of results obtained by different methods of individual dietary survey. Br J Nutr 2:89–110.

Dufour DL, Staten LK, Reina JC, Spurr GB. 1997. Living on the edge: dietary strategies of economically impoverished women in Cali, Colombia. Am J Phys Anthropol 102:5–15.

Gersovitz M, Madden JP, Smiciklas-Wright H. 1978. Validity of the 24-hr. dietary recall and seven-day record for group comparisons. J Am Diet Assoc 73: 48–55.

Greger JL, Etnyre GM. 1978. Validity of 24-hour dietary recalls by adolescent females. Am J Public Health 68:70–72.

Hankin JH, Rhoads GG, Gobler GA. 1975. A dietary method for an epidemiologic study of gastrointestinal cancer. Am J Clin Nutr 28:1055–1061.

Huss-Ashmore R. 1996. Issues in the measurement of energy intake for free-living human populations. Am J Hum Biol 8:159–168.

Instituto Nacional de Nutrición. 1988. Tabla de composición de alimentos Colombianos, 5th ed. Bogota, Colombia: Instituto Nacional de Nutrición.
Linusson EEI, Sanjur D, Erickson EC. 1974. Validating the 24-hour recall method as a dietary survey tool. Arch Latinoam Nutr 24:277–294.
Lohman TG, Roche AF, Martorell R, editors. 1988. Anthropometric standardization reference manual. Champaign, IL: Human Kinetics.
Madden PJ, Goodman SJ, Guthrie HA. 1976. Validity of the 24-hr. recall. J Am Diet Assoc 68:143–147.
Ohlson MA, Jackson L, Boek J, Cederquist DC, Brewer WD, Brown CG. 1950. Nutrition and dietary habits of aging women. Am J Public Health 40:1101–1108.
Pennington JAT, Church HN. 1985. Food values of portions commonly used, 14th ed. New York: Harper & Row Publishers.
Romieu I, Hernandez-Avila M, Rivera JA, Ruel MT, Parra S. 1997. Dietary studies in countries experiencing a health transition: Mexico and Central America. Am J Clin Nutr 65:1159S–1165S.
Rutishauser IHE. 1982. Food models, photographs or household measures? Proc Nutr Soc Aust 7:144.
Spurr GB, Dufour DL, Reina JC. 1996. Energy expenditure of urban Colombian women: a comparison of patterns and total daily energy expenditure by the heart rate and factorial methods. Am J Clin Nutr 63:870–878.
Thomas RU, Fox HM, Kelly HJ, Moyer EZ, Macy IG. 1954. Rapid method for qualitative appraisal of food intake of groups. J Am Diet Assoc 30:865–871.
Thomson AM. 1958. Diet in pregnancy. Br J Nutr 12:446–461.
Thompson FE, Byers T. 1994. Dietary assessment resource manual. J Nutr 124:2245S–2317S.
USDA. 1994. USDA nutrient data bank (online data base).
Young CM, Hagan GC, Tucker RE, Foster WD. 1952. A comparison of dietary study methods. II Dietary history vs. seven-day record vs. 24hr recall. J Am Diet Assoc 28:218–221.